Korean J Leg Med 2014;38:48-58
http://dx.doi.org/10.7580/kjlm.2014.38.2.48

원 저

차세대 염기서열 분석법을 이용한 15개 상염색체 STR의 염기서열 생성 및 유전자형 분석

김은혜[1,2] · 정상은[1] · 신경진[1,2]
양우익[1] · 양인석[1]

[1]연세대학교 의과대학 법의학과
[2]연세대학교 BK21 플러스
연세의과학사업단

접 수 : 2014년 4월 25일
수 정 : 2014년 5월 9일
게재승인 : 2014년 5월 13일

본 연구과제는 2012년도 대검찰청의 '범죄자 DNA DB 및 DNA 감식 기술의 국산화 및 차세대 선진기술 기반 구축 연구개발비(1333-304-260)'의 지원을 받아 수행되었습니다.

책임저자 : 양인석
(120-752) 서울시 서대문구 연세로 50, 연세대학교 의과대학 법의학과
전화 : +82-2-2228-2691
FAX : +82-2-362-0860
E-mail : graduate@nate.com

Sequence Generation and Genotyping of 15 Autosomal STR Markers Using Next Generation Sequencing

Eun Hye Kim[1,2], Sang-Eun Jung[1], Kyoung-Jin Shin[1,2], Woo Ick Yang[1], In Seok Yang[1]

[1]Department of Forensic Medicine, Yonsei University College of Medicine, Seoul, Korea
[2]Brain Korea 21 PLUS Project for Medical Science, Yonsei University, Seoul, Korea

Recently, next generation sequencing (NGS) has received attention as the ultimate genotyping method to overcome the limitations of capillary electrophoresis (CE)-based short tandem repeat (STR) analysis, such as the limited number of STR loci that can be measured simultaneously using fluorescent-labeled primers and the maximum size of STR amplicons. In this study, we analyzed 15 autosomal STR markers via the NGS method and evaluated their effectiveness in STR analysis. Using male and female standard DNA as single-sources and their 1:1 mixture, we sequentially generated sample amplicons by the multiplex polymerase chain reaction (PCR) method, constructed DNA libraries by ligation of adapters with a multiplex identifier (MID), and sequenced DNA using the Roche GS Junior Platform. Sequencing data for each sample were analyzed via alignment with pre-built reference sequences. Most STR alleles could be determined by applying a coverage threshold of 20% for the two single-sources and 10% for the 1:1 mixture. The structure of the STR in each allele was accurately determined by examining the sequences of the target STR region. The mixture ratio of the mixed sample was estimated by analyzing the coverage ratios between assigned alleles at each locus and the reference/variant ratios from the observed sequence variations. In conclusion, the experimental method used in this study allowed the successful generation of NGS data. In addition, the NGS data analysis protocol enables accurate STR allele call and repeat structure determination at each locus. Therefore, this approach using the NGS system will be helpful to interpret and analysis the STR profiles from singe-source and even mixed samples in forensic investigation.

Key Words : Short tandem repeat, Next generation sequencing, Repeat structure, Sequence variation, Mixture

서 론

법과학 분야에서 주로 사용되는 짧은연쇄반복(short tandem repeat; 이하 STR)은 사람의 유전체(genome)의 비암호화 영역(non-coding region)에 존재하며 이는 2~7 base pair (bp)의 염기서열이 반복적으로 나타나는 특징을 가진다. STR은 개인마다 핵심반복단위(core repeat unit)의 반복수

pISSN 1225-0589
eISSN 2287-2078

(repeat number)가 다르게 나타나고 개인마다 고유한 값을 가지기 때문에 개인식별과 혈연관계의 확인 목적으로 STR 분석을 활용하고 있다.[1-3]

현재 법과학 실무에서는 중합효소연쇄반응(polymerase chain reaction; 이하 PCR)으로 얻은 증폭 산물을 모세관 전기영동법(capillary electrophoresis; 이하 CE)으로 분리하여 길이의 차이에 따른 STR의 반복 수를 조사하여 STR 유전자형을 분석하고 있다.[3] 이때 여러 STR 유전좌에 대한 증폭 산물이 동시에 얻어질 수 있도록 다중증폭 PCR (multiplex PCR)이 많이 이용된다. CE 기반의 분석법은 단 한 개의 염기 차이도 구별이 가능한 해상도를 갖고 있어 증폭 산물의 길이를 정확하게 확인할 수 있으며, 형광표지자가 부착된 시동체(primer; 프라이머)를 이용하여 증폭 산물을 자동화된 장비에서 쉽고 빠르게 검출할 수 있다. 그러나 이 방법은 증폭 산물의 염기서열을 확인할 수 없을 뿐 아니라 사용할 수 있는 형광표지자의 수 및 증폭 산물의 크기에서 제한이 있다.

기존의 염기서열을 분석하는 방법인 Sanger 기반의 염기서열 분석법은 정확하게 염기서열 정보를 얻을 수는 있지만, 개인 유전체 분석(personal genome analysis) 등 대용량의 DNA 염기서열 정보를 얻어야 하는 연구 분야에 적용하는 것은 분석에 걸리는 시간, 노동력, 비용 측면에서 비효율적이다. 이 때문에 고효율과 저비용으로 대용량의 DNA 염기서열 정보를 얻을 수 있는 새로운 분석기법에 대한 요구가 있었다. 2000년대 중반에 주형 DNA를 대상으로 짧은 길이의 염기서열을 대용량으로 빠르게 생성시킬 수 있는 차세대 염기서열 분석법(next generation sequencing; NGS)이 소개되었다.[4] NGS 장비의 개발과 시약의 개선이 이루어지고, 생물정보학적 기법이 발달함에 따라서 NGS 분석은 기존의 Sanger 기반의 방법을 대체할 수 있는 여러 가지 장점을 가지고 있어 많은 연구 분야에서 사용되고 있다.[5-9]

법과학 분야에서도 새로운 NGS 기법을 STR 분석에 적용해 봄으로써 기존의 CE 기반의 방법과 비교하여 어떠한 장점을 가지고 있는지, 특히 기존의 방법에서 나타나는 STR 분석의 제한점이 극복될 수 있는지 알아보는 시도가 이루어져 왔으며, 최근 이에 대한 연구결과들이 속속 발표되고 있다.[10-15] 하지만 NGS 기법으로 STR 증폭 산물의 염기서열 분석을 위한 시료 준비 및 라이브러리 제작과 같은 실험적 방법과 생성된 NGS 자료로부터 STR 대립유전자형을 결정하는 분석법이 아직 확고하게 확립되지 않았다. 따라서 본 연구에서는 STR 분석에 주로 사용되고 있는 다중증폭 PCR 방법으로 얻어진 증폭 산물로부터 NGS 자료를 생성하기 위한 최적의 실험적 방법과 생성된 대용량의 NGS 자료의 분석을 통해 STR 대립유전자형 결정, 대립유전자의 반복구조, 염기서열변이를 효과적으로 분석하는 방법을 제시함으로써 단일시료뿐만 아니라 1:1 혼합시료에 대해서도 함께 NGS를 이용한 STR 유전자형 분석의 유용성을 알아보고자 한다.

재료 및 방법

1. DNA 시료

사용된 DNA 시료는 법의유전학 연구에서 대조군으로 사용되고 있는 상용 남성 표준 시료 2800M (Promega, Madison, WI, USA), 여성은 9947A (Promega)를 사용하였다. 이들 DNA 시료는 NanoDrop 1000 spectrophotometer (Thermo. Fisher scientific, Waltham, MA, USA)를 이용하여 정량한 후 1 ng/㎕의 농도로 준비하였다. 1:1 혼합시료는 두 개의 단일시료(2800M과 9947A)를 섞어서 최종농도 1 ng/㎕가 되도록 했다.

2. STR 증폭 산물의 생성 및 확인

본 연구에서는 D3S1358, TH01, D21S11, D18S51, Penta E, D5S818, D13S317, D7S820, D16S539, CSF1PO, Penta D, vWA, D8S1179, TPOX, FGA의 15개 STR 유전좌 및 Amelogenin을 분석할 수 있도록 PowerPlex 16 system (Promega)의 공개된 정보를 바탕으로 프라이머(primer)를 준비하였다. 다만 기존의 CE를 이용한 STR 분석법과 다르게 프라이머에 형광표지자를 부착하지 않았다. Table 1은 PCR 과정에서 사용된 프라이머의 서열 및 최종농도를 보여준다. PCR은 2.5 ㎕의 10X Gold ST*R buffer (Promega), 4.0 unit의 AmpliTaq Gold DNA polymerase (Applied Biosystems, Foster City, CA, USA), primer와 1 ng의 DNA 시료를 포함하는 총 25 ㎕의 반응액을 준비하여 PowerPlex 16 system에서 권장하는 방법에서 PCR 온도 순환만 34회로 조정하여 수행하였다. PCR을 마친 후에 폴리아크릴아마이드젤 전기영동(polyacrylamide gel electrophoresis)을 통해서 증폭 산물들이 균일하게 생성되었는지 확인하였다. 생성된 증폭 산물의 정제는 QIAquick® PCR purification Kit (QIAGEN, Hilden, Germany)를 이용하였다. 얻어진 증폭 산물의 농도는 Quant-iT™ PicoGreen® dsDNA Assay Kit (Invitrogen, Carlsbad, CA, USA)를 이용하여 측정했으며, 순도 측정은 NanoDrop 1000 spectrophotometer (Thermo. Fisher scientific)로 260 nm와 280 nm의 파장에서 측정된 흡광도의 비율을 계산함으로써 이루어졌다.

3. 증폭 산물을 이용한 라이브러리(library)의 제작

NGS 분석을 위한 첫 단계로 생성된 증폭 산물을 대상으로 특정 어댑터(adapter)를 붙여주는 라이브러리 제작은 GS

Rapid Library Preparation Kit (Roche Diagnostics Corp., Branford, CT, USA)를 이용하여 제조사의 지시대로 수행하였다. 이 과정에서 DNA 시료에 따른 구분을 위해서 Multiplex Identifier (MID)가 포함된 어댑터를 사용하였다. 제작된 라이브러리의 정제는 AMPure bead (Beckman Coulter, Brea, CA, USA)를 이용하였는데, 증폭 산물과 비드(bead)의 비율이 2:1이 되도록 함으로써 크기가 100 bp 미만인 작은 절편들을 제거할 수 있도록 하였다. 최종적으로 얻어진 라이브러리에 대한 크기별 분포 확인 및 농도 측정은 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA)를 이용하였다.

4. 클론 증폭(clonal amplification) 및 대량 염기서열 생성

제작된 라이브러리의 클론 증폭을 위한 에멀션 PCR (emulsion PCR)은 GS Junior Titanium emPCR Kit (Lib-L; Roche Diagnostics Corp.)을 이용하여 제조사의 지시대로 수행하였다. 이를 위해서 라이브러리는 측정된 농도를 기반으로 1 ㎕ 당 분자의 수를 계산한 후에 1 ㎕ 당 1×10^7개의 분자가 되도록 희석하였다(Eq. 1). 또한 Scheible 등[14]이 제시한대로 copy per bead 수는 1.0으로 설정하였다. 에멀션 PCR 증폭 산물에 대한 대량 염기서열 생성은 GS Junior Titanium Sequencing Kit (Roche Diagnostics Corp.)를 이용하여 제조사의 지시에 따라 GS Junior (Roche Diagnostics Corp.) 장비에서 수행하였다.

$$Molecules/\mu l = \frac{(sample\ conc.;\ ng/\mu l)\times 6.022\times 10^{23}}{656.6\times(Average\ amplicon\ length;bp)} \quad \text{Eq. 1}$$

5. NGS 자료의 분석

STR 대립유전자형 결정을 위해 Bornman 등[15]이 제시한 프로토콜에 기초하여 i) 참조서열(reference sequence)의 제작, ii) NGS 리드와 참조서열 간의 정렬(alignment), iii) 각 STR

Table 1. Adjusted Final Concentrations of Primer Sets for Multiplex PCR system*

Loci	Primer	Primer sequences (5' →3')	Final Conc. (uM)
D3S1358	D3-PP16-F	ACTGCAGTCCAATCTGGGT	0.20
	D3-PP16-R	ATGAAATCAACAGAGGCTTGC	
TH01	TH01-PP16-F	GTGATTCCCATTGGCCTGTTC	0.10
	TH01-PP16-R	ATTCCTGTGGGCTGAAAAGCTC	
D21S11	D21-PP16-F	ATATGTGAGTCAATTCCCCAAG	0.60
	D21-PP16-R	TGTATTAGTCAATGTTCTCCAGAGAC	
D18S51	D18-PP16-F	TTCTTGAGCCCAGAAGGTTA	0.50
	D18-PP16-R	ATTCTACCAGCAACAACACAAATAAAC	
Penta E	PentaE-PP16-F	ATTACCAACATGAAAGGGTACCAATA	1.20
	PentaE-PP16-R	TGGGTTATTAATTGAGAAAACTCCTTACAATTT	
D5S818	D5-PP16-F	GGTGATTTTCCTCTTTGGTATCC	0.20
	D5-PP16-R	AGCCACAGTTTACAACATTTGTATCT	
D13S317	D13-PP16-F	ATTACAGAAGTCTGGGATGTGGAGGA	0.40
	D13-PP16-R	GGCAGCCCAAAAAGACAGA	
D7S820	D7-PP16-F	ATGTTGGTCAGGCTGACTATG	0.30
	D7-PP16-R	GATTCCACATTTATCCTCATTGAC	
D16S539	D16-PP16-F	GGGGGTCTAAGAGCTTGTAAAAAG	0.40
	D16-PP16-R	GTTTGTGTGTGCATCTGTAAGCATGTATC	
CSF1PO	CSF1PO-PP16-F	CCGGAGGTAAAGGTGTCTTAAAGT	0.30
	CSF1PO-PP16-R	ATTTCCTGTGTCAGACCCTGTT	
Penta D	PentaD-PP16-F	GAAGGTCGAAGCTGAAGTG	1.20
	PentaD-PP16-R	ATTAGAATTCTTTAATCTGGACACAAG	
Amelogenin	Amelo-PP16-F	CCCTGGGCTCTGTAAAGAA	0.25
	Amelo-PP16-R	ATCAGAGCTTAAACTGGGAAGCTG	
vWA	vWA-PP16-F	GCCCTAGTGGATGATAAGAATAATCAGTATGTG	0.15
	vWA-PP16-R	GGACAGATGATAAATACATAGGATGGATGG	
D8S1179	D8-PP16-F	ATTGCAACTTATATGTATTTTTGTATTTCATG	0.50
	D8-PP16-R	ACCAAATTGTGTTCATGAGTATAGTTTC	
TPOX	TPOX-PP16-F	GCACAGAACAGGCACTTAGG	0.15
	TPOX-PP16-R	CGCTCAAACGTGAGGTTG	
FGA	FGA-PP16-F	GGCTGCAGGGCATAACATTA	0.60
	FGA-PP16-R	ATTCTATGACTTTGCGCTTCAGGA	

*Each primer sequence based on the information from PowerPlex 16 system without fluorescent dye

http://dx.doi.org/10.7580/kjlm.2014.38.2.48

대립유전자에서의 coverage 값의 계산, iv) 각 STR 유전좌에서의 대립유전자형 결정의 순서로 분석이 이루어졌다. 참조서열의 제작을 위해 현재까지 알려진 STR 대립유전자의 반복수 및 이들의 서열은 STRbase (http://www.cstl.nist.gov/biotech/strbase)로부터 얻었으며, 각 STR의 5′ 및 3′ 주변부 서열(flanking region sequence)은 human genome GRCh37/hg19에서 가져왔다. 또한, 주변부 서열의 길이를 500~550 bp로 설정함으로써 어떠한 primer 조합을 통해 얻은 NGS 리드도 참조서열과의 정렬(alignment)이 이루어질 수 있도록 하였다(Fig. 1). 최종적으로 참조서열은 5′ 주변부 서열, STR 영역의 서열, 3′ 주변부 서열로 구성될 수 있도록 Microsoft® Excel®의 매크로 기능을 이용하여 제작하였다. NGS 리드와 참조서열간의 정렬은 리눅스(Linux) 운영체제에서 Bowtie 2[16] 프로그램을 이용하였다. 얻어진 결과 파일의 형식전환을 위해 SAMtools[17]과 BEDTool[18]을 순차적으로 사용하였다. 각 STR 대립유전자에 대한 coverage 값은 참조서열에 정렬된 리드 중에서 전체 STR 영역을 포함하는 리드의 수를 계산함으로써 얻어졌다. 각 STR 유전좌에서 대립유전자형의 결정을 위해서 단일시료에서는 각 유전좌에서 전체 coverage 값의 20%, 혼합시료에서는 10%로 기준값을 적용함으로써 이루어졌다. 앞에서 결정된 각 STR 대립유전자의 염기서열을 바탕으로 한 반복구조(repeat structure)의 확인은 Integrative Genomics Viewer[19]를 이용하였다. 또한, 각 STR 유전좌에서 대립유전자형에 대한 coverage 값의 비율을 조사하는 방법과 특정 위치에서 나타나는 염기서열변이(sequence variation)를 확인하고, 각 염기의 비율을 알아보는 방법으로

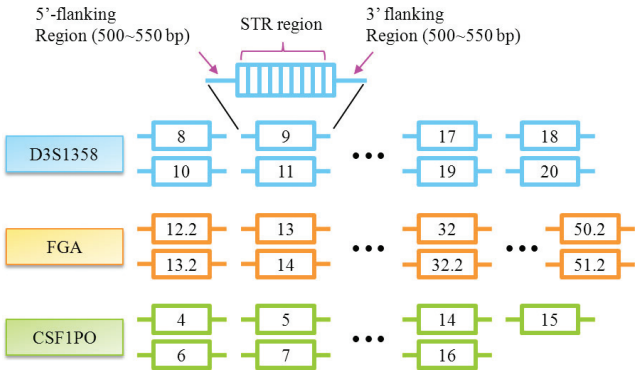
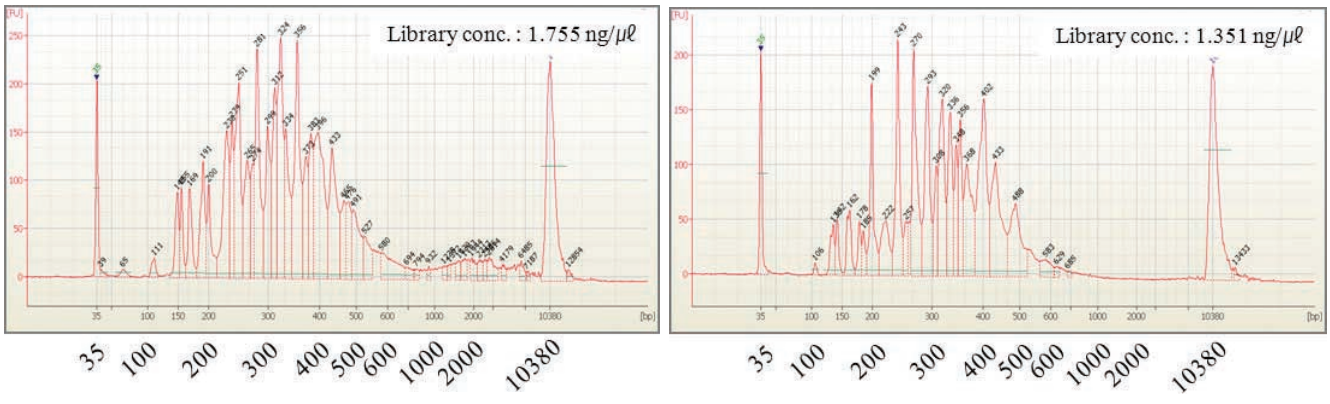


Fig. 1. Schematic view of STR reference sequences. Long flanking sequences ranged between 500 bp and 550 bp in STR reference sequences were designed for complete alignment of sample sequences that generated with any primer combinations.



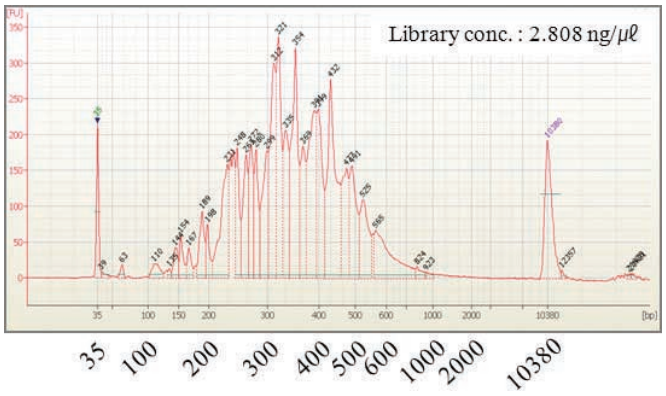


Fig. 2. Quality check of constructed libraries on High Sensitivity chip using 2100 Bioanalyzer. Fragments less than 100 bp including adaptor dimers were successfully removed.
a: Standard male DNA 2800M ; b: Standard female DNA 9947A ; c: 1:1 mixture

1:1 혼합시료 비율을 추정하였다.

6. 모세관 전기영동(capillary electrophoresis)을 이용한 STR 분석

NGS 자료로부터 얻어진 남녀 표준시료 및 1:1 혼합시료의 대립유전자형이 정확하게 결정되었는지 확인하기 위해서 모세관 전기영동 기반의 STR 분석법으로 이들 시료의 대립유전자형을 알아보았다. 이를 위해 각 DNA 시료 1 ng과 PowerPlex 16 HS system (Promega)을 이용하여 제조사의 지시대로 PCR을 수행하고, 얻어진 증폭 산물은 ABI PRISM 3130xl Genetic Analyzer와 GeneScan Software Version 3.7 (Applied Biosystems)을 이용하여 검출하였으며, 마지막으로 GeneMapper™ ID Software Version 3.1 (Applied Biosystems)을 이용하여 분석하였다.

결 과

1. PCR 증폭 산물로부터 라이브러리의 제작

Powerplex 16 system 정보를 바탕으로 준비된 프라이머를 이용하여 2800M 표준 남성 DNA 시료, 9947A 표준 여성 DNA 시료, 이들의 1:1 혼합시료로부터 PCR을 수행하였으며, 얻어진 증폭 산물을 정제 한 후에 이들의 순도를 측정했을 때 1.92~1.95의 범위로 나왔으며 1650~2220 ng의 범위로 증폭 산물을 얻었다. 따라서 Roche사에서 제시하는 라이브러리 제작을 위한 최소량 500 ng 이상, 순도 1.70 이상 2.00 미만의 기준을 충족하였기에 3개 DNA 시료의 증폭 산물로부터 라이브러리를 제작하였다. Fig. 2는 Bioanalyzer를 통해 얻은 최종적으로 얻은 각 시료의 라이브러리에 대한 크기별 분포를 보여주고 있다. 여기서 100 bp 미만의 작은 절편들은 거의 확인되지 않기 때문에 라이브러리 제작과정에서 비드를 이용한 방법으로 작은 절편들이 선택적으로 제거됨을 확인할 수 있었다.

2. NGS 자료의 시료 별 분류 및 서열정렬(sequence alignment)

NGS을 통해 얻은 리드의 수가 총 164,468개였으며, 이들의 평균길이는 183.64 bp로 나왔으며, MID 서열을 이용한 시료에 따른 분류를 통해서 2800M 표준시료는 51,475개, 9947A 표준시료는 33,213개, 이들의 1:1 혼합시료는 76,943개, 그리고 분류되지 않은 리드는 2,837개로 얻어졌다. 이들 자료를 참조서열과의 정렬을 통해 각 STR 유전좌에서 얻어진 리드 수를 확인할 수 있었다(Table 2). 15개의 STR 유전좌 중 D3S1358, D5S818, D13S317, TH01에서는 다른 유전좌들에 비해 많은 리드 수가 얻어졌다. 그리고 D16S439, D18S51, CSF1PO, FGA, Penta D, Penta E, TPOX에서는 상대적으로 적은 리드를 얻었는데, 이들 증폭 산물의 크기는 대체로 250

Table 2. Read Counts of 15 STR Loci in Each Sample

STR locus	Amplicon size range (bp)	2800M			9947A			1:1 mixture		
		All*	Entire STR†	Entire STR/ All (%)	All	Entire STR	Entire STR/ All (%)	All	Entire STR	Entire STR/ All (%)
D3S1358	115-147	9470	8743	92.3	6341	6012	94.8	14261	13306	93.3
D5S818	119-155	9485	8705	91.8	5523	5011	90.7	9347	8531	91.3
D7S820	215-247	3676	3476	94.6	1868	1780	95.3	4815	4603	95.6
D8S1179	203-247	4458	4017	90.1	1967	1805	91.8	3368	3054	90.7
D13S317	169-201	4897	4631	94.6	4060	3868	95.3	12839	12140	94.6
D16S439	264-304	967	877	90.7	708	655	92.5	2497	2361	94.6
D18S51	209-366	739	332	**44.9**	1284	546	**42.5**	1117	481	**43.1**
D21S11	203-259	3045	2313	76.0	2996	2525	84.3	4873	3871	79.4
CSF1PO	321-357	291	244	83.8	596	522	87.6	862	742	86.1
FGA	322-444	956	460	**48.1**	666	255	**38.3**	3137	1440	**45.9**
Penta D	376-441	142	31	**21.8**	267	56	**21.0**	403	75	**18.6**
Penta E	379-474	193	84	**43.5**	356	116	**32.6**	563	309	**54.9**
TH01	156-195	5503	4620	84.0	3324	2811	84.6	6712	5518	82.2
TPOX	262-290	269	230	85.5	215	183	85.1	679	576	84.8
vWA	123-171	3153	2782	88.2	1014	919	90.6	8565	7649	89.3
AMEL	106, 112	3416	3247	95.1	1773	1741	98.2	2334	2247	96.3
Total		50550	44792	88.6	32958	28805	87.4	76372	66903	87.6

*All aligned reads regardless of the presence or absence of STR region
†Aligned reads containing entire STR region
Entire STR with less than 50% represents in bold text

http://dx.doi.org/10.7580/kjlm.2014.38.2.48

bp보다 컸음을 확인할 수 있었다. 각 유전좌에서 모든 리드의 수에 대한 전체 STR 영역을 포함하는 리드의 수의 백분율을 조사하였을 때 D18S51, FGA, Penta D, Penta E에서 50% 미만으로 나오는 것을 볼 수 있었다. 마찬가지로 증폭 산물의 크기가 클수록 전체 STR 영역을 포함하는 리드의 수도 적게 얻어졌음을 보여준다.

3. STR 대립유전자의 반복구조 결정 및 염기서열변이의 확인

Table 3은 2개의 단일시료(2800M 및 9947A)와 이들의 1:1 혼합물에 대한 NGS 자료로부터 STR 대립유전자가 결정되는 예시를 보여준다. 이와 같은 방법으로 이들 시료에 대해서 15개 STR 유전좌에서의 대립유전자를 결정할 수 있었다(Table 4). NGS 자료로부터 결정된 유전자형이 정확하게 일치하는지 확인하기 위하여 기존 CE 분석법으로 결정된 STR 유전자형 결과와 비교해 본 결과 단일시료의 경우에는 15개 STR 유전좌에서 모두 일치하였고, 1:1 혼합시료의 경우 13개 STR 유전좌(D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11, CSF1PO, FGA, Penta D, TH01, TPOX)에서는 대립유전자형이 정확하게 일치하였으나, 2개 유전좌(Penta E, vWA)에서는 일치하지 않았다. 이들의 coverage 값을 토대로 확인해 본 결과 Penta E의 대립유전자 12는 7.44%, vWA의 대립유전자 18은 9.11%로 혼합시료의 분석 기준값인 10%에 미치지 못하였다.

남녀 표준 DNA 시료의 NGS 자료로부터 15개 STR 유전좌

Table 3. Determination of D3S1358 Alleles based on Percentage of Allele Coverage in 2 Single-sources and 1:1 Mixture

Alleles	2800M		9947A		1:1 mixture	
	Allele read count	Allele coverage* (%)	Allele read count	Allele coverage* (%)	Allele read count	Allele coverage* (%)
11	0		2	0.03	0	
12	0		12	0.20	5	0.04
13	2	0.02	217	3.61	103	0.77
14	13	0.15	2868	47.70	1519	11.41
15	71	0.81	2879	47.89	2245	16.87
16	495	5.66	34	0.57	541	4.06
17	4355	49.81	0		4936	37.09
18	3757	42.97	0		3906	29.35
19	24	0.27	0		33	0.25
20	26	0.30	0		21	0.16
Total	8743		6012		13309	

Shaded sections indicate assigned alleles based on the analytical threshold
*Percentage of allele coverage (%) = allele read count/locus read count × 100

Table 4. STR Genotyping Results in 2 Single-sources and 1:1 Mixture examined by CE and NGS Analyses

STR locus	2800M		9947A		1:1 mixture	
	CE	NGS	CE	NGS	CE	NGS
D3S1358	17, 18	17, 18	14, 15	14, 15	14, 15, 17, 18	14, 15, 17, 18
D5S818	12	12	11	11	11, 12	11, 12
D7S820	8, 11	8, 11	10, 11	10, 11	8, 10, 11	8, 10, 11
D8S1179	14, 15	14, 15	13	13	13, 14, 15	13, 14, 15
D13S317	9, 11	9, 11	11	11	9, 11	9, 11
D16S539	9, 13	9, 13	11, 12	11, 12	9, 11, 12, 13	9, 11, 12, 13
D18S51	16, 18	16, 18	15, 19	15, 19	15, 16, 18, 19	15, 16, 18, 19
D21S11	29, 31.2	29, 31.2	30	30	29, 30, 31.2	29, 30, 31.2
CSF1PO	12	12	10, 12	10, 12	10, 12	10, 12
FGA	20, 23	20, 23	23, 24	23, 24	20, 23, 24	20, 23, 24
Penta D	12, 13	12, 13	12	12	12, 13	12, 13
Penta E	7, 14	7, 14	12, 13	12, 13	7, 12, 13, 14	7, (12), 13, 14
TH01	6, 9.3	6, 9.3	8, 9.3	8, 9.3	6, 8, 9.3	6, 8, 9.3
TPOX	11	11	8	8	8, 11	8, 11
vWA	16, 19	16, 19	17, 18	17, 18	16, 17, 18, 19	16, 17, (18), 19

Alleles in parentheses represent true allele with coverage value less than 10% of total coverage value

Table 5. Repeat Structures of 15 STRs in Two Standard Samples from NGS Data

A. 2800M

STR locus	Genotype	Core repeat	Repeat structure
D3S1358	17, 18	TCTA	17: TCTA $[TCTG]_3$ $[TCTA]_{13}$ 18: TCTA $[TCTG]_3$ $[TCTA]_{14}$
D5S818	12	AGAT	12: $[AGAT]_{12}$
D7S820	8, 11	GATA	8: $[GATA]_8$ 11: $[GATA]_{11}$
D8S1179	14, 15	TCTA	14: TCTA TCTG $[TCTA]_{12}$ 15: $[TCTA]_2$ TCTG $[TCTA]_{12}$
D13S317	9, 11	TATC	9: $[TATC]_9$ $\underline{[AATC]_2}$ 11: $[TATC]_{11}$ $\underline{TATC}$ $\underline{AATC}$
D16S539	9, 13	GATA	9: $[GATA]_9$ 13: $[GATA]_{13}$
D18S51	16, 18	AGAA	16: $[AGAA]_{16}$ $\underline{AAAG}$ $\underline{[AG]_3}$ 18: $[AGAA]_{18}$ $\underline{AAAG}$ $\underline{[AG]_3}$
D21S11	29, 31.2	TCTA	29: $[TCTA]_4$ $[TCTG]_6$ $[TCTA]_3$ TA $[TCTA]_3$ TCA $[TCTA]_2$ TCCA TA $[TCTA]_{11}$ 31.2: $[TCTA]_5$ $[TCTG]_6$ $[TCTA]_3$ TA $[TCTA]_3$ TCA $[TCTA]_2$ TCCA TA $[TCTA]_{11}$ TA TCTA
CSF1PO	12	AGAT	12: $[AGAT]_{12}$
FGA	20, 23	CTTT	20: $[TTTC]_3$ TTTT TTCT $[CTTT]_{12}$ CTCC $[TTCC]_2$ 23: $[TTTC]_3$ TTTT TTCT $[CTTT]_{15}$ CTCC $[TTCC]_2$
Penta D	12, 13	AAAGA	12: $[AAAGA]_{12}$ 13: $[AAAGA]_{13}$
Penta E	7, 14	AAAGA	7: $[AAAGA]_7$ 14: $[AAAGA]_{14}$
TH01	6, 9.3	AATG	6: $[AATG]_6$ 9.3: $[AATG]_6$ ATG $[AATG]_3$
TPOX	11	AATG	11: $[AATG]_{11}$
vWA	16, 19	TCTA	16: TCTA $[TCTG]_3$ $[TCTA]_{12}$ TCCA TCTA 19: TCTA $[TCTG]_4$ $[TCTA]_{14}$ TCCA TCTA

B. 9947A

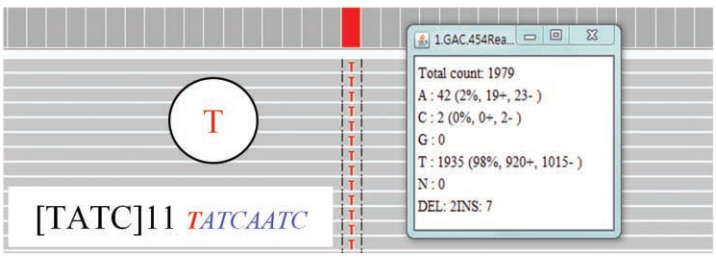
STR locus	Genotype	Core repeat	Repeat structure
D3S1358	14, 15	TCTA	14: TCTA $[TCTG]_2$ $[TCTA]_{11}$ 15: TCTA $[TCTG]_2$ $[TCTA]_{12}$
D5S818	11	AGAT	12: $[AGAT]_{11}$
D7S820	10, 11	GATA	10: $[GATA]_{10}$ 11: $[GATA]_{11}$
D8S1179	13	TCTA	13a: TCTA TCTG $[TCTA]_{11}$ 13b: $[TCTA]_{13}$
D13S317	11	TATC	11: $[TATC]_{11}$ $\underline{[AATC]_2}$
D16S539	11, 12	GATA	11: $[GATA]_{11}$ 12: $[GATA]_{12}$
D18S51	15, 19	AGAA	15: $[AGAA]_{15}$ $\underline{AAAG}$ $\underline{[AG]_3}$ 19: $[AGAA]_{19}$ $\underline{AAAG}$ $\underline{[AG]_3}$
D21S11	30	TCTA	30: $[TCTA]_6$ $[TCTG]_5$ $[TCTA]_3$ TA $[TCTA]_3$ TCA $[TCTA]_2$ TCCA TA $[TCTA]_{11}$
CSF1PO	10, 12	AGAT	10: $[AGAT]_{10}$ 12: $[AGAT]_{12}$
FGA	23, 24	CTTT	23: $[TTTC]_3$ TTTT TTCT $[CTTT]_{15}$ CTCC $[TTCC]_2$ 24: $[TTTC]_3$ TTTT TTCT $[CTTT]_{16}$ CTCC $[TTCC]_2$
Penta D	12	AAAGA	12: $[AAAGA]_{12}$
Penta E	12, 13	AAAGA	12: $[AAAGA]_{12}$ 13: $[AAAGA]_{13}$
TH01	8, 9.3	AATG	8: $[AATG]_8$ 9.3: $[AATG]_6$ ATG $[AATG]_3$
TPOX	8	AATG	8: $[AATG]_8$
vWA	17, 18	TCTA	17: TCTA $[TCTG]_4$ $[TCTA]_{12}$ TCCA TCTA 18: TCTA $[TCTG]_4$ $[TCTA]_{13}$ TCCA TCTA

http://dx.doi.org/10.7580/kjlm.2014.38.2.48

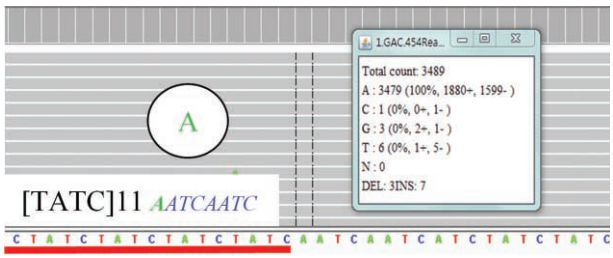
에서 결정된 대립유전자의 염기서열을 확인하였으며, 이를 바탕으로 각 STR 영역의 반복구조를 결정할 수 있었다(Table 5). 또한, 각 시료 간에 STR 유전좌에서의 염기서열을 비교하여 다음과 같이 반복구조의 차이 혹은 염기서열의 변이를 관찰하였다. 첫 번째는 두 개의 대립유전자형이 길이는 같지만, 염기서열이 다른 경우이다. 9947A 시료의 D8S1179 유전좌는 CE 기반의 분석법으로는 대립유전자형이 13, 13으로 동형접합자(homozygous)로 나타나지만, NGS를 통해 분석한 결과 하나는 "TCTA TCTG $[TCTA]_{11}$"으로, 다른 하나는 "$[TCTA]_{13}$"으로 서로 다른 반복구조를 가진 대립유전자형으로 나타나는 것으로 확인되었다. 결과적으로는 STR 영역의 길이는 갖지만, 서로 다른 염기서열을 갖는 이형접합자(heterozygous)인 것이다. 두 번째는 시료 간에 서로 다른 반복구조를 가진 경우이다. D3S1358 유전좌에서는 핵심반복단위는 [TCTA]로 시료에 따라 [TCTG]의 반복단위가 발견된다. 2800M과 9947A 시료 간에 D3S1358 유전좌의 반복구조를 비교했을 때 [TCTG]가 나타나는 위치가 각각 세 번째와 두 번째로 서로 다르게 나타나는 것이 관찰되었다. 세 번째로 STR 영역이 아닌 주변부 서열에서 염기서열변이가 관찰된 경우이다. STR 대립유전자형은 9, 11로 확인된 2800M 시료의 D13S317 유전좌에서 유일하게 관찰되었다. 이들의 3' 주변부 서열에서 대립유전자 9는 "AATC AATC"로, 대립유전자 11은 "TATC AATC"로 나타났다. 마치 [TATC]의 반복이 하나 더 추가된 것처럼 관찰된 것이다.

4. 혼합시료에서의 혼합비율 추정

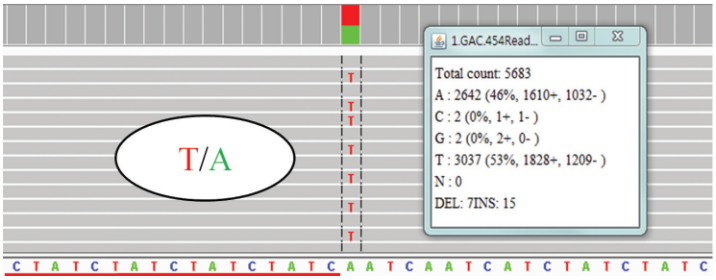
분석 대상인 15개의 각 STR 유전좌에서 대립유전자형에 대한 coverage 값의 비율을 조사하는 방법과 특정 위치에서 나타나는 서열변이를 확인하고 각 염기의 비율을 알아보는 방법으로 1:1 혼합시료 비율을 추정하였다. D3S1358 유전좌를 예로 들면, 2800M에서는 "17, 18"의 대립유전자형을 가지고 있고, 9947A에서는 "14, 15"이기 때문에 이들의 1:1 혼합물에 대한 대립유전자형은 "14, 15, 17, 18"이 된다. 이론적으로 각 대립유전자형에 대한 coverage 값의 비율이 1:1:1:1로 예상되었으나, 이들의 coverage 값이 각각 1519, 2245, 4936, 3006의 순으로 나와서(Table 3) 이들의 비율은 "1:1.5:3.3:2.6"으로 얻어졌다. 또한, 2800M 시료의 D13S317 유전좌에서는 대립유전자 11의 3´ 주변부 서열에서 human reference genome hg19을 기준으로 아데닌(adenine)에서 티민(thymine)으로의 염기서열변이가 확인되었다(Fig. 3). 이 위치에서 각 염기의 수를 조사한 결과 전체 coverage 값 5683 중에서 티민은 3037 (46%)로, 아데닌은 2642 (53%)으로 나와서 두 개의 염기가 거의 1:1로 존재하고 있음을 확인하였다. 그런데 2800M 시료의 D13S317 유전좌에서의 대립유전자형은 9, 11의 이형접합자이고 9947A는 11, 11의 동형접합자라는 점을 감안한다면 아데닌과 티민의 비율이 2:1로 나와야만 한다. 결국, 예상되는 실제 혼합비율과 다르게 나타났다는 것을 알 수 있었다.



a



b



c

Sample (coverage)	A	T	G	C
1:1 mixture (5683)	2642 (46%)	3037 (53%)	2	2

d

Fig. 3. Estimation of mixture ratio based on reference/variant ratios from observed sequence variations in D13S317 locus. The sequence variation of adenine (A) to thymine (T) was detected in 3´ flanking region of D13S317 locus. Mixture ratio was estimated to 46% (A) : 53% (T).
a: Standard male DNA 2800M ; b: Standard female DNA 9947A ; c: 1:1 mixture ; d: Mixture ratio

고 찰

NGS 방법을 이용한 STR 유전자형 분석은 STR 증폭 산물의 생성, 라이브러리 제작, 대량의 염기서열 생성, 자료 분석의 과정으로 이루어진다. 본 연구에서는 법과학 영역에서 CE 기반의 상염색체 STR 분석에 사용되는 PowerPlex 16 system의 프라이머 정보를 이용하여 다중증폭 PCR 시스템을 구축하고 표준 DNA 시료를 대상으로 증폭 산물을 생성하고 NGS를 통해 얻은 결과를 분석하였다. 본 연구는 여러 그룹에서도 보고한 법과학 STR을 대상으로 NGS 기법으로 분석한 방법과 비슷하다.[10-15] Van Neste 등의 첫 번째 연구[10]에서는 9개의 STR 유전좌를 분석할 수 있는 상용화된 STR 키트를 이용하여 단일시료 및 혼합시료에 대한 증폭 산물을 준비하고 NGS 분석을 수행하였다. 여기서는 형광표지자가 부착된 프라이머를 그대로 사용하였는데, 저자들은 NGS 분석 결과로 정방향(forward) 및 역방향(reverse)으로 읽은 리드 수의 차이가 크게 나타나는 점을 확인했으며, 이것이 형광표지자의 영향일 것으로 추측하였다. 이 때문에 본 연구에서도 Van Neste 등의 두 번째 연구[20]와 동일하게 15개 STR 유전좌에 대해 형광표지자가 부착되지 않은 프라이머를 가지고 다중증폭 PCR 방법으로 증폭 산물을 준비하고 다만 다른 NGS 장비인 GS Junior를 사용하여 분석하였다.

라이브러리 제작을 위해 Roche 사에서 권장하는 방법은 i) 어댑터 서열과 주형 특이적 서열(template specific sequence)이 서로 결합된 프라이머(퓨전 프라이머; fusion primer)를 이용하여 증폭 산물을 생성하거나, ii) 온전한 주형 DNA를 작은 절편으로 만드는 과정(절편화; fragmentation)을 수행한 후 어댑터를 붙이는 방법이다. 첫 번째로 퓨전 프라이머를 이용하는 방법은 이전에 증폭 산물 생성과 라이브러리 제작을 동시에 진행하기 위해 사용해 본 적이 있다. 하지만 증폭 산물을 전반적으로 고르게 얻지 못하였고, 이에 따라 NGS로 얻어진 총 리드 수가 적게 나왔으며, 일부 유전좌에서 대립유전자형이 정확하게 결정되지 않은 경우가 발생하였다. 아마도 길어진 프라이머가 사용되었기 때문에 PCR 과정에서의 증폭 효율이 떨어졌고, 또한 이어 진행된 에멀션 PCR 단계에서도 영향을 끼친 것으로 생각되어 본 연구에서는 사용되지 않았다. 두 번째로 DNA를 절편화 하는 방법은 전장 유전체(whole genome) 및 미토콘드리아 DNA와 같이 길이가 긴 경우에 라이브러리를 제작하는 방법으로 다중증폭 PCR을 통해서 100~450 bp 범위의 증폭 산물을 생성함으로써 이루어지는 STR 분석에는 적절하지 않았다. 하지만 이러한 증폭 산물을 이미 절편화가 완료된 작은 절편으로 간주하고 어댑터를 부착을 통한 라이브러리 제작을 통해서 성공적으로 NGS 자료를 생성할 수 있었다. 결과적으로 기존의 다중증폭 PCR의 방식을 그대로 유지하면서 위와 같이 라이브러리를 제작하는 방법은 법과학 분야에서 NGS를 통한 STR 유전좌의 연구에 매우 유용할 것이라고 본다.

15개 STR 유전좌에 대해서 NGS 자료를 생성하고, 분석을 통해 각 유전좌마다 리드의 분포를 조사했을 때 일정하게 나오지 않고 증폭 산물의 크기와 반비례하여 나타나는 것이 관찰되었다(Table 2). 대체로 250 bp를 기준으로 이것보다 증폭 산물이 작게 만들어지는 유전좌에서는 리드의 수가 많게 나왔지만, 크게 나오는 유전좌에 대해서는 리드 수가 상대적으로 적게 얻어졌다. 특히 300 bp 이상의 증폭 산물이 생성되는 D18S51, FGA, Penta D, Penta E에서는 모든 리드(All)의 수도 적게 얻어졌을 뿐만 아니라 이들 중에서 전체 STR 영역을 포함하는 리드의 비율(Entire STR/All)도 50% 미만으로 확인되었다. 본 연구에서는 NGS를 위해 PowerPlex 16 system의 프라이머 정보를 이용하였기 때문에 이에 따른 증폭 산물의 크기도 106~474 bp의 범위로 넓게 나타나게 된다. 이러한 점들을 고려할 때 전체적으로 증폭 산물의 크기를 줄이면서 보다 좁은 범위에서 이들이 생성될 수 있게 한다면, 각 STR 유전좌마다 일정한 리드의 수를 얻게 됨으로써 차후 분석결과에 신뢰를 줄 수 있을 것으로 예상된다. 따라서 NGS에 최적화된 STR 분석결과를 얻기 위해서는 새로운 실험적 설계가 필요할 것으로 본다. 또한, GS Junior 장비 이외에 다른 시퀀싱 방식을 사용하는 동급의 MiSeq (Illumina Inc., San Diego, CA, USA) 및 Ion Torrent PGM (Life Technologies, Carlsbad, CA, USA) 장비에서도 성능 개선을 통해 읽을 수 있는 리드의 길이가 점차 길어지고 있기 때문에 이러한 장비에서도 함께 적용될 수 있는 설계가 요구될 것이다.

단일시료 및 1:1 혼합시료를 NGS를 통해 STR 대립유전자형을 결정한 후에 CE 분석법으로 얻어진 결과와 비교하였을 때, 단일시료에서는 모든 STR 유전좌에서 대립유전자형이 일치하였는데 반하여 1:1 혼합시료의 일부 STR 유전좌(Penta E, vWA)에서는 CE 분석법으로 얻은 대립유전자형과 NGS 분석으로 얻은 대립유전자형이 서로 일치하지 않는 것이 확인되었다(Table 4). 이것은 이들 유전좌에서 각 하나씩의 대립유전자의 coverage 값이 본 연구에서 대립유전자 결정을 위해 설정한 기준값(10%) 미만으로 나왔기 때문이다. 그렇지만 이들 대립유전자에서는 stutter라고 여겨지는 대립유전자의 coverage 값보다는 크게 나왔기 때문에 결과에서 이들을 배제하는 것은 옳지 않다고 판단하였다. 앞으로도 NGS를 이용한 혼합시료의 STR 분석에서도 대립유전자를 결정할 때 위와 같은 점을 고려하여 분석 결과에 오류가 없도록 세심한 노력이 필요할 것으로 본다.

NGS 기법으로 STR 대립유전자의 반복구조 결정 및 염기서열변이의 관찰이 가능하여(Table 5), 또한 두 개의 남녀 표준시료(2800M과 9947A)에서 3가지의 특징을 확인할 수 있었다. 첫 번째는 한 유전좌에서 같은 길이의 대립유전자로 보였

http://dx.doi.org/10.7580/kjlm.2014.38.2.48

지만 다른 염기서열을 갖고 있는 경우였으며, 두 번째는 한 유전좌에서 서로 다른 시료 간에 다른 반복구조를 보이는 경우였고, 세 번째는 STR 영역의 반복구조는 같지만, 주변부 서열에서 염기서열변이가 관찰된 경우였다. 이러한 점들은 NGS를 이용한 염기서열 기반의 분석으로 기존의 CE를 통해 확인된 STR 대립유전자가 더욱 더 세분될 수 있음을 시사한다. 또한, 앞선 연구[11]에서 제시한 바와 같이, 한국인에서도 NGS를 이용한 STR 대립유전자의 염기서열정보 및 이들의 빈도자료가 구축된다면 앞으로 친자확인 및 범죄수사와 같은 법과학 실무에 유용할 것이다.

1:1 혼합시료에서 NGS 분석을 통해 혼합비율의 추정하기 위해 STR 대립유전자에 대한 coverage 값의 비율로 알아보는 방법을 사용하였을 때 얻어진 결과 값이 예상하고 있는 비율과 다르게 나오는 것이 확인되었다. 특이하게도 2800M과 9947A에서 각각 유래된 대립유전자를 분리하여 coverage 값을 조사하였을 때 동일한 양상으로 나오지 않고, 2800M 유래의 대립유전자 쪽으로 치우치는 경향을 확인할 수 있었다(Table 3). 이러한 양상은 15개 STR 유전좌에서 모두 동일하게 나타났다(자료 제시 없음). 뿐만 아니라 D13S317 유전좌에서 관찰된 염기서열변이로부터 아데닌과 티민의 수를 조사하여 혼합비율을 추정한 경우에서도 2800M에서 유래된 티민이 예상보다 많이 나오는 것이 관찰되었다(Fig. 3). 이러한 원인을 알아보기 위해서 CE를 통해 얻은 1:1 혼합시료의 프로필(profile)에서 대립유전자의 피크(peak) 높이를 조사하여 혼합비율을 추정해보았다. CE 결과에서도 NGS 결과와 마찬가지로 한쪽 시료의 대립유전자가 예상보다 많이 나온다는 것을 알 수 있었다(자료 제시 없음). CE 및 NGS 기법은 공통적으로 대상 시료로부터 PCR을 통해서 증폭 산물을 준비하는 것으로 시작한다. 이것으로 볼 때 위와 같은 현상은 PCR 과정에서 발생하는 두 개의 시료 간의 증폭 효율의 차이라는 것을 미루어 짐작할 수 있었다. 따라서 NGS를 이용하여 혼합비율을 추정하는 경우에는 이러한 점을 충분히 고려하여 분석이 이루어져야 할 것이다.

본 연구 및 Bornman 등[15]의 연구에서는 2개의 단일시료를 이용하여 "1:1"의 비율에 대해서만 NGS를 통한 분석을 수행하였다. 이러한 경우는 용의자와 피해자가 각각 한 명으로 구성된 사건 현장에서 얻어진 시료를 해석하는데 적용될 수 있을 것이다. 하지만 혹시라도 둘 중 한 명의 시료에서 낮은 비율로 나타난다면 자료의 해석이 어려워질 수 있다. 따라서 사건 현장에서 얻어지는 시료의 실제적인 특성을 고려하기 위해서는 "1:1" 조건 이외에도 좀 더 다양한 비율로 혼합된 시료를 대상으로 효과적인 자료 해석이 이루어지는지 조사할 필요가 있다. Van Neste 등은 총 4개의 시료로부터 "10:20:30:40" 및 "93.40:5:1:0.5:0.1"의 비율로 혼합시료를 만들어 NGS를 통한 분석에 이용하였다.[20] 여기서 분석에 사용된 최소 기준을 0.5%로 설정하였기 때문에 이론적으로는 1%로 존재하는 시료까지는 검출되어야 하지만, 실제적으로는 5% 이상으로 존재하는 시료부터 검출할 수 있었다. 이러한 연구는 NGS를 이용한 혼합물 분석에서 가장 큰 관심거리인 "소수의 공여자(minor contributor)로부터의 대립유전자를 얼마나 낮은 비율까지(민감도; sensitivity) 그리고 얼마나 정확하게(특이도; specificity) 검출할 수 있는가"를 알아보는 데 중요한 정보를 제공할 것으로 본다. 앞으로 이러한 연구 결과를 NGS 자료의 분석을 통해 STR 대립유전자형을 결정하는 데 활용함으로써 얻어진 자료의 해석이 정확하게 이루어질 수 있도록 노력해야 할 것으로 본다.

본 연구에서 제시한 NGS 자료 분석의 전략으로 참조서열을 직접 제작함으로써 법과학에서 주로 사용되는 STR 유전좌에 맞게 대립유전자형을 결정할 수 있도록 새로운 방법을 제시하였다. 이전에 다른 연구자들에 의해 개발된 STR을 분석하는 lobSTR 프로그램도 보고된 바 있다.[21] 하지만 본 연구에서 제시한 분석법이 lobSTR 프로그램을 사용했을 때보다 향상된 결과를 보였다(자료 제시 없음). 본 연구에서 사용된 분석법은 복잡하고 번거로운 과정 때문에 실제 사용자들이 느끼기에는 다소 어려운 점이 있을 것으로 여겨진다. 이에 새로 제작된 참조서열을 이용한 분석 프로그램이 개발된다면 좀 더 효율적으로 NGS 분석을 수행할 수 있을 것이고 더 나아가 좀 더 많은 STR 분석의 적용에도 유용할 것으로 생각한다.

본 연구에서는 남녀 표준시료 단일시료 및 이들의 혼합시료를 대상으로 단 한 번의 NGS 과정을 통해 성공적으로 염기서열 자료를 생성할 수 있었을 뿐만 아니라 이들 자료로부터 효과적인 STR 분석을 수행할 수 있었다. 이러한 방법은 범죄현장에서 발견된 시료와 함께 용의자 및 피해자에게서 채취한 시료의 분석에 대해서도 동일하게 적용될 수 있는 모델이라 판단된다. 따라서 NGS를 이용한 STR 분석법이 실험적, 분석적 측면에서 보다 최적화가 이루어진다면 기존의 CE 기반의 방법이 가지는 부족한 점을 채워줌으로써 법과학 분야에서 기존 방법과 함께 추가적인 방법으로 유용하게 사용 될 수 있을 것으로 전망한다.

Acknowledgment

본 연구를 위해서 GS Junior 장비 사용에 도움을 주신 대검찰청 DNA 수사담당관실 담당자 여러분께 감사드립니다.

참 고 문 헌

1. Thompson R, Zoppis S, McCord B. An overview of DNA typing methods for human identification: past, present, and future. Methods Mol Biol 2012;830:3-16.
2. Kayser M, de Knijff P. Improving human forensics

through advances in genetics, genomics and molecular biology. Nat Rev Genet 2011;12:179-92.
3. Berglund EC, Kiialainen A, Syvänen AC. Next-generation sequencing technologies and applications for human genetic history and forensics. Investig Genet 2011;2:23.
4. Metzker ML. Sequencing technologies - the next generation. Nat Rev Genet 2010;11:31-46.
5. Cho IS, Blaser MJ. The human microbiome: at the interface of health and disease. Nat Rev Genet 2012;13:260-70.
6. Bamshad MJ, Ng SB, Bigham AW, et al. Exome sequencing as a tool for Mendelian disease gene discovery. Nat Rev Genet 2012;12:745-55.
7. Ozsolak F, Milos PM. RNA sequencing: advances, challenges and opportunities. Nat Rev Genet 2011;12:87-98.
8. Meyerson M, Gabriel S, Getz G. Advances in understanding cancer genomes through second-generation sequencing. Nat Rev Genet 2010;11:685-96.
9. Laird PW. Principles and challenges of genomewide DNA methylation analysis. Nat Rev Genet 2010;11:191-203.
10. Van Neste C, Van Nieuwerburgh F, Van Hoofstat D, et al. Forensic STR analysis using massive parallel sequencing. Forensic Sci Int Genet 2012;6:810-8.
11. Rockenbauer E, Hansen S, Mikkelsen M, et al. Characterization of mutations and sequence variants in the D21S11 locus by next generation sequencing. Forensic Sci Int Genet 2014;8:68-72.
12. Fordyce SL, Ávila-Arcos MC, Rockenbauer E, et al. High-throughput sequencing of core STR loci for forensic genetic investigations using the Roche Genome Sequencer FLX platform. Biotechniques 2011;51:127-33.
13. Dalsgaard S, Rockenbauer E, Buchard A, et al. Non-uniform phenotyping of D12S391 resolved by second generation sequencing. Forensic Sci Int Genet 2014;8:195-9.
14. Scheible M, Loreille O, Just R, et al. Short tandem repeat sequencing on the 454 platforms. Forensic Sci Int Genet Suppl Ser 2011;3:357-8.
15. Bornman DM, Hester ME, Schuetter JM, et al. Short-read, high-throughput sequencing technology for STR genotyping. Biotechniques 2012;0:1-6.
16. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods 2012;9:357-9.
17. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics 2009;25:2078-9.
18. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 2010;26:841-2.
19. Robinson JT, Thorvaldsdóttir H, Winckler W, et al. Integrative genomics viewer. Nat Biotechnol 2011;29:24-6.
20. Van Neste C, Vandewoestyne M, Van Criekinge W, et al. My-Forensic-Loci-queries (MyFLq) framework for analysis of forensic STR data generated by massive parallel sequencing. Forensic Sci Int Genet 2014;9:1-8.
21. Gymrek M, Golan D, Rosset S, et al. lobSTR: a short tandem repeat profiler for personal genomes. Genome Res 2012;22:1154-62.

 http://dx.doi.org/10.7580/kjlm.2014.38.2.48